Network of Cancer Genes (NCG 3.0): integration and analysis of genetic and network properties of cancer genes

Matteo D'Antonio, Vera Pendino, Shruti Sinha and Francesca D. Ciccarelli*

Department of Experimental Oncology, European Institute of Oncology, IFOM-IEO Campus, Via Adamello 16, 20139 Milan, Italy

ABSTRACT

The identification of a constantly increasing number of genes whose mutations are causally implicated in tumor initiation and progression (cancer genes) requires the development of tools to store and analyze them. The Network of Cancer Genes (NCG 3.0) collects information on 1494 cancer genes that have been found mutated in 16 different cancer types. These genes were collected from the Cancer Gene Census as well as from 18 whole exome and 11 whole-genome screenings of cancer samples. For each cancer gene, NCG 3.0 provides a summary of the gene features and the cross-reference to other databases. In addition, it describes duplicability, evolutionary origin, orthology, network properties, interaction partners, microRNA regulation and functional roles of cancer genes and of all genes that are related to them. This integrated network of information can be used to better characterize cancer genes in the context of the system in which they act. The data can also be used to identify novel candidates that share the same properties of known cancer genes and may therefore play a similar role in cancer. NCG 3.0 is freely available at http://bio.ifom-ieo-campus.it/ncg.

INTRODUCTION

The pivotal role of genomic instability in causing human cancer is an old concept that dates back to the first cytogenetic studies on cancer cells (1,2). Until very recently, however, the number of genes whose somatic mutations are causally implicated in cancer initiation and progression was very low. This was mainly due to the difficult and lengthy process of identifying genetic modifications through traditional sequencing methods. The recent development of next-generation sequencing techniques has boosted the discovery of novel cancer genes. In the last 5 years, high-throughput mutational screenings were performed on several cancer types and led to the identification of mutations in both coding and non-coding regions of the cancer genome. So far, around 30 high-throughput and whole genome sequencing experiments on cancer samples have been published, which identified around 1500 mutated genes that are potentially actively involved in cancer development (3–30). This constant delivery of new sequencing data is radically changing our understanding of cancer genetics, showing that the genomic landscape of cancer is complex and varies among and within cancer types (31). One way to reduce the complexity implies the identification of properties that are shared among cancer genes and distinguish them from the rest of human genes (32). For example, cancer genes tend to be singletons, i.e. they preserve one gene copy in the genome, and to encode hubs, i.e. proteins that engage numerous physical connections with other proteins (33). Moreover, recessive cancer genes appeared early in evolution and are involved in basic cellular processes, while oncogenes originated later and mainly act as regulators and signal transducers (34). Thus, the identification of specific properties of cancer genes may help in better rationalizing their role in cancer. While several databases collect mutational and functional information on cancer genes, no description of their systems-level properties is currently available. The Network of Cancer Genes (NCG) was originally created to collect information on duplicability and network properties of cancer genes (35).

Information on duplicability and network properties of cancer genes (35). In the current version of the database (NCG 3.0), we included 1494 cancer genes that have been identified so far, and provide an update of the human gene set and the protein interactome. We also added novel features, such as the functional description of cancer genes as well as information about their interaction with microRNAs (miRNAs). Finally, we developed a new web

---

*To whom correspondence should be addressed. Tel: +39 02574303053; Fax: +39 0294375990; Email: francesca.ciccarelli@ifom-ieo-campus.it

Table 1. Primary data correspond to the list of cancer genes as they were extracted from the original papers. Five genes are not present in NCG 3.0 because they could not be mapped to current version of Entrez IDs. Dominant and recessive genes refer to the definition reported in the cancer gene census (37). Ancient and recent genes refer to genes that originated between the last universal common ancestor and opisthokonts and with metazoans or later, respectively. HTMS, high-throughput mutation screening; WGS, whole genome sequencing

Cancer genes	Dominant	Recessive	Amplified	HTMS	WGS	Total
Primary data	348	98	77	699	457	1,499
Present in NCG 3	346	96	77	698	454	1,494
Duplicated (%)	58 (17.7)	10 (10.5)	23 (29.9)	102 (14.8)	64 (14.3)	230 (15.7)
Ancient (%)	184 (55.6)	63 (66.3)	53 (69.9)	377 (54.6)	273 (61.2)	837 (57.1)
Recent (%)	147 (44.4)	32 (33.7)	24 (31.1)	313 (45.4)	173 (38.8)	629 (42.9)
Hubs (%)	185 (58.6)	71 (75.5)	50 (70.4)	231 (42.6)	116 (34.6)	532 (44.7)
miRNA targets (%)	53 (15.3)	14 (14.6)	17 (22.1)	53 (7.6)	28 (6.2)	118 (7.9)
miRNA hosts (%)	12 (3.5)	6 (6.3)	2 (2.6)	28 (4.0)	17 (3.7)	55 (3.7)

interface to allow more intuitive and flexible queries of the database. Since properties of cancer genes can be also exploited to identify new candidates that may be involved in cancer (36), we present a case study that involves two paralogs, the DNA methyltransferase alpha (*DNMT3A*) and beta (*DNMT3B*). On the basis of the information that can be retrieved in NCG 3.0, we hypothesize that, upon the acquisition of somatic mutations, these genes may play an active role in cancer development.

UPDATES OF DATA SOURCES

Cancer genes

NCG 3.0 includes information on 1494 cancer genes that were extracted from 30 distinct studies (Table 1). The genes were divided into three groups, according to the experimental evidence that supported their involvement in cancer. The first group includes 498 genes from the latest release of the Cancer Gene Census (CGC, 444 genes, 31 March 2011) (37) and from the census of amplified cancer genes (77 genes) (38). Both these resources are literature-based collections of genes whose mutations and/or amplifications have been proven to be causally implicated in tumorigenesis. Among the 444 genes from CGC, 346 are defined as dominant (i.e. heterozygous mutations are sufficient to cause tumorigenesis) and 96 as recessive (i.e. homozygous mutations are required to initiate tumorigenesis) (37). The second group is formed of 698 candidate cancer genes from 18 high-throughput mutational screenings (HTMS) on 16 distinct cancer types, such as breast (15,30), colorectal (30), pancreatic (13–15), lung (10,15), renal (8), prostatic (15), bladder (12), and ovarian (15) cancers, glioblastoma (20,21), leukemia (11), lymphoma (23), sarcoma (4), myeloma (6), medulloblastoma (22), head and neck squamous cell carcinoma (3,27), and melanoma (29). These genes represent the subsets of mutated genes that were identified as bearing driver mutations, i.e. mutations that confer growth advantage and are actively involved in tumor development (31). The third group of cancer genes contains 457 genes that have been identified in eleven whole genome sequencing (WGS) screenings of cancer genomes, such as breast (26), lung (16,25), prostatic (5), liver cancer (28), glioblastoma (7), leukemia (17,19), myeloma (6) and melanoma (24).

Human gene set and orthology information

We gathered information on the human gene set from the combination of Gencode v.7.0 (39) and RefSeq v. 46 (40). Gencode is an initiative that aims at identifying and mapping all human protein coding genes (41). It has been used as the reference gene set for the Encode project, for the 1000 Genomes Project and for the design of the capture baits for whole exome sequencing (39). This gene set is therefore likely to include all genes present in current and future mutational screenings of whole cancer exomes. Starting from 20 700 genes in Gencode v.7.0 (corresponding to 84 408 protein sequences), we removed multiple isoforms that map to the same gene locus by aligning all protein sequences to the reference human genome (hg18) and retaining only the longest isoform (33). At the end of this pipeline, we were able to retrieve 19 560 unique genes. The 1140 missing genes corresponded to transcripts that spanned more than one gene. To recover them, we repeated the same pipeline with 20 750 genes in RefSeq (corresponding to 34 571 protein sequences). The union of these two gene lists led to the final dataset of 20 531 unique human genes.

To measure the duplicability of each gene, we identified all additional hits on the genome that span at least 60% of the gene length (33), and found that 4311 human genes (21% of the total) have at least one additional copy. Consistently with previous reports (33,34), we found that cancer genes duplicate significantly less than the rest of human genes (15.7% of the total, p-value $= 3.0\,10^{-7}$, Pearson's Chi-squared Test), and dominant genes are more duplicated than recessive genes (Figure 1A and Table 1).

We updated the information on orthology relationships between human and other species to eggNOG v. 2.0 (42). Orthologs were used to identify the evolutionary origin of 1,466 cancer genes (98% of the total), defined as the deepest internal node of the tree of life where an ortholog could be found (34). We confirmed that recessive cancer genes are mostly ancient genes that have orthologs in prokaryotes, plants and fungi, while a higher fraction of
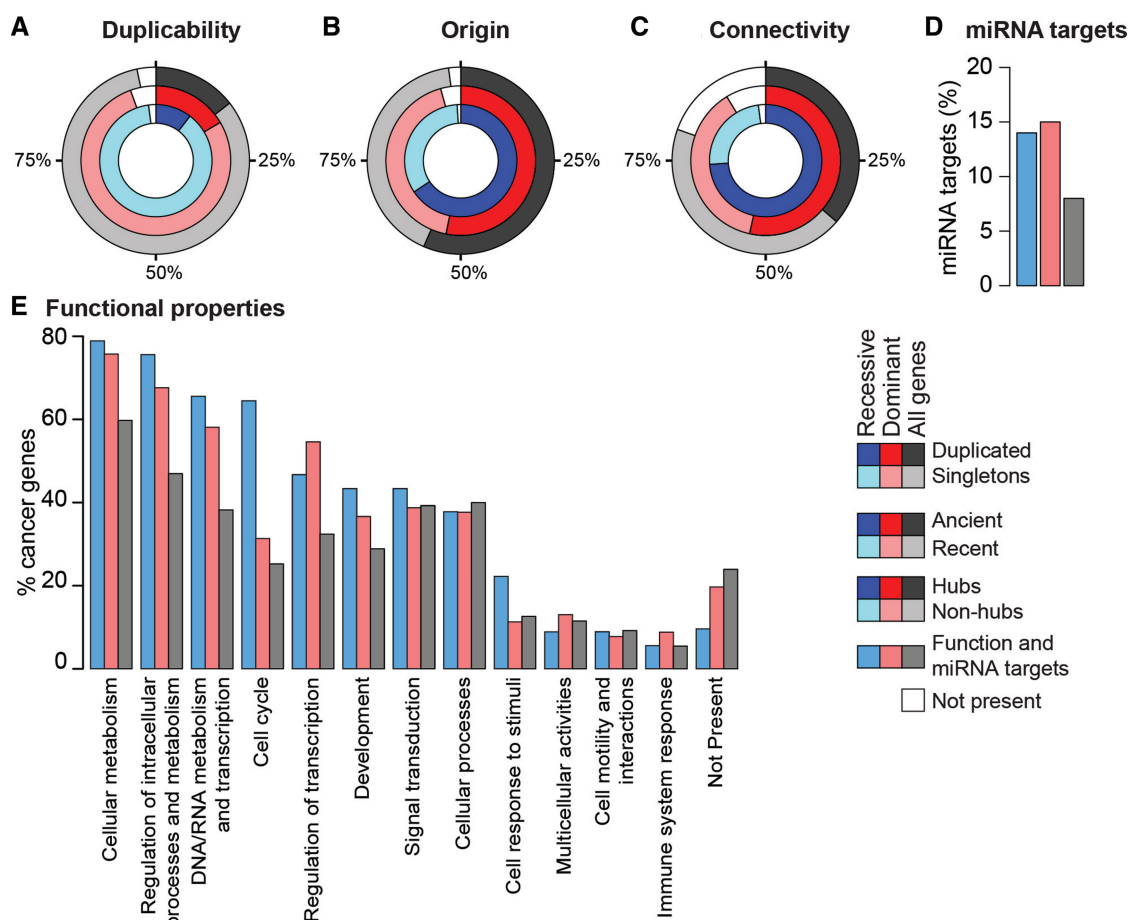
Figure 1. Properties of cancer genes. Circles represent the fraction of cancer, dominant, and recessive genes that are (**A**) singleton or duplicated, (**B**) ancient or recent, (**C**) hubs or non-hubs. Ancient genes originated between the last universal common ancestor and opisthokonts; recent genes originated with metazoans or later. (**D**) miRNA targets were derived from TarBase (50) and miRecords (51). Approximately 15% of dominant and recessive cancer genes are targets of miRNAs, compared with 8% of all cancer genes. (**E**) Cancer genes were associated to one of the 12 functional categories based on the corresponding GO terms (34).

dominant cancer genes appeared with metazoans or later (Figure 1B and Table 1).
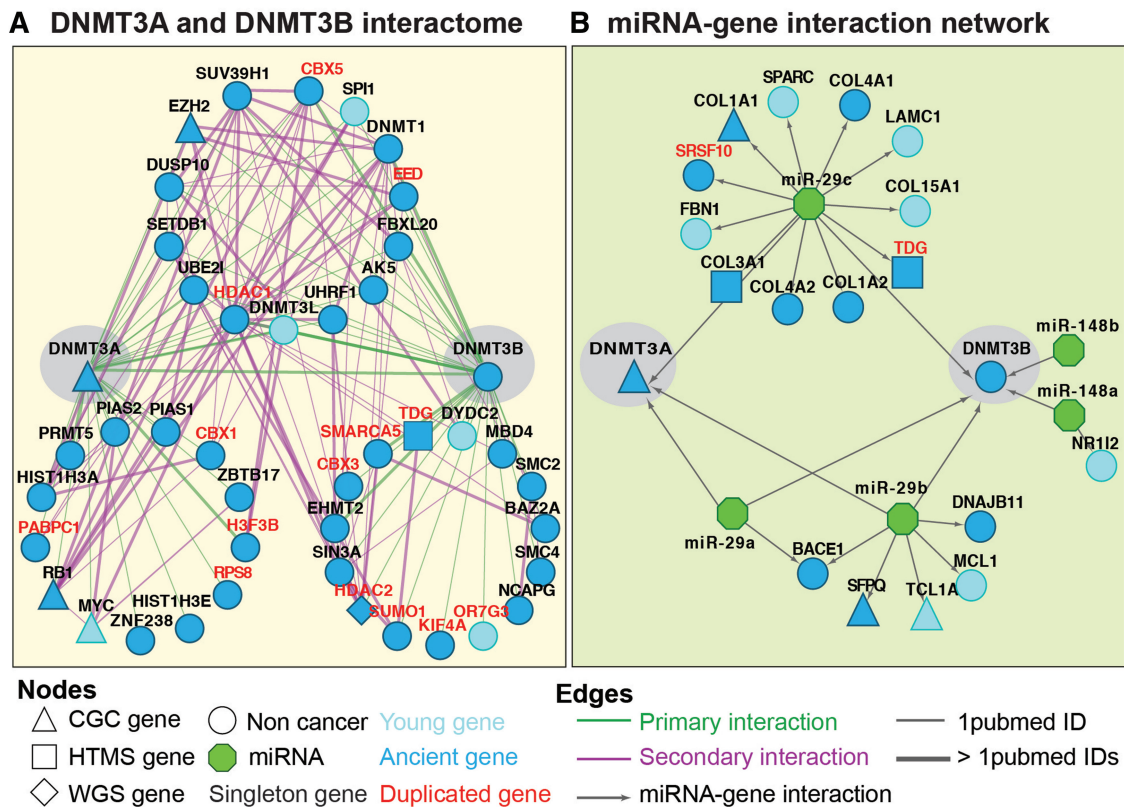
Human interaction network

We rebuilt the human interactome by integrating the latest releases of five sources of protein–protein interactions, such as HPRD (43), BioGRID (44), IntAct (45), MINT (46) and DIP (47). Only primary interactions were used while putative interactions inferred from orthology relationships were discarded. The final dataset comprised 98 492 binary interactions between 13 531 proteins, supported by 25 915 independent publications. The resulting human interactome contains 13% more proteins and 44% more interactions than the previous version of NCG, and reports interaction data for 1189 cancer genes (79% of the total). We confirmed that cancer genes are enriched in hubs, defined as top 25% most connected proteins (at least 15 interactions) (34), and are therefore more connected than the rest of human genes (Figure 1C and Table 1).

NEW FEATURES OF NCG 3.0

Interactions between cancer gene and miRNAs

miRNAs are endogenous short nucleotide sequences that interfere with RNA transcripts in the cytoplasm, thus regulating protein synthesis (48). Alterations in miRNA expression have been reported during cancer initiation, progression and metastasis (49). In addition, miRNAs regulate well-known cancer genes, such as PTEN, NRAS, KRAS, (48). Because of this involvement in cancer, we included information about the interactions between miRNAs and their target cancer genes. Primary data were extracted from TarBase v.5.0 (50) and miRecords v.1.0 (51), which collect experimentally supported interactions between miRNAs and their targets. The non-redundant integration of the two databases returned a list of 54 miRNAs that regulate 118 cancer genes (8% of the total, Table 1). This fraction is significantly higher compared to the rest of human genes (~4%, p-value $= 10^{-11}$, Pearson's Chi-squared Test), and becomes even higher when only recessive or dominant

Figure 2. Functional redundancy of DNMT3A and DNMT3B. (**A**) Protein-protein interaction network of DNMT3A and DNMT3B. The two proteins share 14 interactors, which are mostly involved in the epigenetic control of chromatin and in the regulation of gene expression. Primary interactions are physical interactions of a protein directly with DNMT3A or DNMT3B, while secondary interactions are physical interactions between their primary interactors. (**B**) Interaction network between DNMT3A and DNMT3B and target miRNAs. Both genes are regulated by members of the miR-29 family, whose expression is altered in cancer and is inversely correlated with the gene expression (57). CGC, cancer gene census; HTMS, high-throughput mutational screenings; WGS, whole genome sequencing.

cancer genes are considered (∼15%, Figure 1D and Table 1). This enrichment reflects at least partly the fact that cancer genes are heavily studied and therefore more information are available for them compared to other genes. In addition to miRNA targets, we also retrieved information on 55 cancer genes that host miRNAs within their genomic locus (Table 1).

Functional classification of cancer genes

To derive the gene functional classification, we first extracted 1108 cancer genes (74% of all the total) that have at least one term at levels 5 and 6 of the biological process branch of the Gene Ontology (GO) tree (52). We then grouped all these terms into 12 functional categories, in order to provide a more general description of the gene function (34). Dominant and recessive cancer genes have different functional distributions. In particular, while dominant genes are mostly involved in the regulation of transcription, recessive genes are associated with basic cellular functions such as cell cycle, cell response to stimuli, cellular and DNA/RNA metabolism (Figure 1E).

Network representation

We developed a novel web interface and added several new possibilities of querying the database, thus allowing the user to search for specific cancer genes, for lists of genes and miRNAs related to cancer, and for the presence of cancer genes in genomic regions of interest. In addition, it is now possible to retrieve the interactions between a certain miRNA and the cancer genes that are associated to it, as well as all cancer genes that are miRNA targets or that host miRNAs within their genomic locus. Finally, we adopted CytoscapeWeb (53) for the visualization of protein–protein and miRNA–target interaction networks, and for displaying the evolutionary origin of cancer genes and their orthologs in other species.

IDENTIFICATION OF CANDIDATE CANCER GENES USING NCG

In addition to providing a comprehensive description of the properties of cancer genes, NCG can be used to identify new possible candidates, based on the hypothesis that genes with systems-level properties similar to known cancer genes may also play similar roles in cancer (36). Following this idea, two almost identical paralogs, *GNAQ GNA11*, have recently been associated to the same tumor type (54). Here, we hypothesize that a similar association can be drawn for *DNMT3A* and its paralog *DNMT3B*. Both these methyltransferases cause

de novo methylation during the differentiation of hematopoietic cells and are involved in hematopoietic stem-cell renewal (55). *DNMT3A* is an already known cancer gene, part of the cancer gene census and frequently mutated in leukemia (56). Although its role in tumorigenesis is still unclear, it has been hypothesized that mutations in *DNMT3A* either induce altered gene expression or affect genome stability (56). To date there is no evidence that also *DNMT3B* is mutated in cancer, although NCG 3.0 provides several indications that this might be indeed the case. First, both *DNMT3A* and *DNMT3B* originated with eukaryotes and are associated with the same GO terms, thus indicating functional redundancy. Second, the encoded proteins DNMT3A and DNMT3B share 14 interactors, which, in turn, are connected through secondary interactions (Figure 2A). The resulting highly interconnected module involves key players in epigenetic changes of chromatin, and consequent regulation of gene expression. Third, both genes are targets of three miRNAs of the miR-29 family (Figure 2B). The downregulation of these miRNAs induces aberrant expression of both *DNMT3A* and *DNMT3B* in non-small-cell lung cancer (57) and the upregulation of both paralogs has been observed in several tumor types (58). We hypothesize that the aberrant expression can be related to the acquisition of mutations in the sequence of these genes.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. von Hansemann,D.P. (1890) On primary cancer of the liver. *Berl. klin. Wschr.*, **27**, 353–356.
2. Boveri,T. (1914) Zur Frage der Entstehung Maligner Tumoren. *Jena: Gustav Fischer*, **1**, 1–64.
3. Agrawal,N., Frederick,M.J., Pickering,C.R., Bettegowda,C., Chang,K., Li,R.J., Fakhry,C., Xie,T.X., Zhang,J., Wang,J. *et al.* (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*, **333**, 1154–1157.
4. Barretina,J., Taylor,B.S., Banerji,S., Ramos,A.H., Lagos-Quintana,M., Decarolis,P.L., Shah,K., Socci,N.D., Weir,B.A., Ho,A. *et al.* (2010) Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat. Genet.*, **42**, 715–721.
5. Berger,M.F., Lawrence,M.S., Demichelis,F., Drier,Y., Cibulskis,K., Sivachenko,A.Y., Sboner,A., Esgueva,R., Pflueger,D., Sougnez,C. *et al.* (2011) The genomic complexity of primary human prostate cancer. *Nature*, **470**, 214–220.
6. Chapman,M.A., Lawrence,M.S., Keats,J.J., Cibulskis,K., Sougnez,C., Schinzel,A.C., Harview,C.L., Brunet,J.P., Ahmann,G.J., Adli,M. *et al.* (2011) Initial genome sequencing and analysis of multiple myeloma. *Nature*, **471**, 467–472.
7. Clark,M.J., Homer,N., O'Connor,B.D., Chen,Z., Eskin,A., Lee,H., Merriman,B. and Nelson,S.F. (2010) U87MG decoded: the genomic sequence of a cytogenetically aberrant human cancer cell line. *PLoS Genet.*, **6**, e1000832.
8. Dalgliesh,G.L., Furge,K., Greenman,C., Chen,L., Bignell,G., Butler,A., Davies,H., Edkins,S., Hardy,C., Latimer,C. *et al.* (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature*, **463**, 360–363.
9. Ding,L., Ellis,M.J., Li,S., Larson,D.E., Chen,K., Wallis,J.W., Harris,C.C., McLellan,M.D., Fulton,R.S., Fulton,L.L. *et al.* (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*, **464**, 999–1005.
10. Ding,L., Getz,G., Wheeler,D.A., Mardis,E.R., McLellan,M.D., Cibulskis,K., Sougnez,C., Greulich,H., Muzny,D.M., Morgan,M.B. *et al.* (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, **455**, 1069–1075.
11. Greif,P.A., Eck,S.H., Konstandin,N.P., Benet-Pages,A., Ksienzyk,B., Dufour,A., Vetter,A.T., Popp,H.D., Lorenz-Depiereux,B., Meitinger,T. *et al.* (2011) Identification of recurring tumor-specific somatic mutations in acute myeloid leukemia by transcriptome sequencing. *Leukemia*, **25**, 821–827.
12. Gui,Y., Guo,G., Huang,Y., Hu,X., Tang,A., Gao,S., Wu,R., Chen,C., Li,X., Zhou,L. *et al.* (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.*, **43**, 875–878.
13. Jiao,Y., Shi,C., Edil,B.H., de Wilde,R.F., Klimstra,D.S., Maitra,A., Schulick,R.D., Tang,L.H., Wolfgang,C.L., Choti,M.A. *et al.* (2011) DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*, **331**, 1199–1203.
14. Jones,S., Zhang,X., Parsons,D.W., Lin,J.C., Leary,R.J., Angenendt,P., Mankoo,P., Carter,H., Kamiyama,H., Jimeno,A. *et al.* (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, **321**, 1801–1806.
15. Kan,Z., Jaiswal,B.S., Stinson,J., Janakiraman,V., Bhatt,D., Stern,H.M., Yue,P., Haverty,P.M., Bourgon,R., Zheng,J. *et al.* (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*, **466**, 869–873.
16. Lee,W., Jiang,Z., Liu,J., Haverty,P.M., Guan,Y., Stinson,J., Yue,P., Zhang,Y., Pant,K.P., Bhatt,D. *et al.* (2010) The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*, **465**, 473–477.
17. Ley,T.J., Mardis,E.R., Ding,L., Fulton,B., McLellan,M.D., Chen,K., Dooling,D., Dunford-Shore,B.H., McGrath,S., Hickenbotham,M. *et al.* (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, **456**, 66–72.
18. Li,M., Zhao,H., Zhang,X., Wood,L.D., Anders,R.A., Choti,M.A., Pawlik,T.M., Daniel,H.D., Kannangai,R., Offerhaus,G.J. *et al.* (2011) Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat. Genet.*, **43**, 828–829.
19. Mardis,E.R., Ding,L., Dooling,D.J., Larson,D.E., McLellan,M.D., Chen,K., Koboldt,D.C., Fulton,R.S., Delehaunty,K.D., McGrath,S.D. *et al.* (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.*, **361**, 1058–1066.
20. McLendon,R., Friedman,A., Bigner,D., Van,M.E.G., Brat,D., Mastrogianakis,G., Olson,J., Mikkelsen,T., Lehman,N., Aldape,K. *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
21. Parsons,D.W., Jones,S., Zhang,X., Lin,J.C., Leary,R.J., Angenendt,P., Mankoo,P., Carter,H., Siu,I.M., Gallia,G.L. *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science*, **321**, 1807–1812.
22. Parsons,D.W., Li,M., Zhang,X., Jones,S., Leary,R.J., Lin,J.C., Boca,S.M., Carter,H., Samayoa,J., Bettegowda,C. *et al.* (2010) The genetic landscape of the childhood cancer medulloblastoma. *Science*, **331**, 435–439.
23. Pasqualucci,L., Trifonov,V., Fabbri,G., Ma,J., Rossi,D., Chiarenza,A., Wells,V.A., Grunn,A., Messina,M., Elliot,O. *et al.* (2011) Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.*

24. Pleasance,E.D., Cheetham,R.K., Stephens,P.J., McBride,D.J., Humphray,S.J., Greenman,C.D., Varela,I., Lin,M.L., Ordonez,G.R., Bignell,G.R. *et al.* (2010) A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, **463**, 191–196.

25. Pleasance,E.D., Stephens,P.J., O'Meara,S., McBride,D.J., Meynert,A., Jones,D., Lin,M.L., Beare,D., Lau,K.W., Greenman,C. *et al.* (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*, **463**, 184–190.

26. Shah,S.P., Morin,R.D., Khattra,J., Prentice,L., Pugh,T., Burleigh,A., Delaney,A., Gelmon,K., Guliany,R., Senz,J. *et al.* (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, **461**, 809–813.

27. Stransky,N., Egloff,A.M., Tward,A.D., Kostic,A.D., Cibulskis,K., Sivachenko,A., Kryukov,G.V., Lawrence,M., Sougnez,C., McKenna,A. *et al.* (2011) The mutational landscape of head and neck squamous cell carcinoma. *Science*, **333**, 1157–1160.

28. Totoki,Y., Tatsuno,K., Yamamoto,S., Arai,Y., Hosoda,F., Ishikawa,S., Tsutsumi,S., Sonoda,K., Totsuka,H., Shirakihara,T. *et al.* (2011) High resolution characterization of a hepatocellular carcinoma genome. *Nat. Genet.*, **43**, 464–469.

29. Wei,X., Walia,V., Lin,J.C., Teer,J.K., Prickett,T.D., Gartner,J., Davis,S., Stemke-Hale,K., Davies,M.A., Gershenwald,J.E. *et al.* (2011) Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat. Genet.*, **43**, 442–446.

30. Wood,L.D., Parsons,D.W., Jones,S., Lin,J., Sjoblom,T., Leary,R.J., Shen,D., Boca,S.M., Barber,T., Ptak,J. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108–1113.

31. Stratton,M.R., Campbell,P.J. and Futreal,P.A. (2009) The cancer genome. *Nature*, **458**, 719–724.

32. Ciccarelli,F.D. (2010) The (r)evolution of cancer genetics. *BMC Biol.*, **8**, 74.

33. Rambaldi,D., Giorgi,F.M., Capuani,F., Ciliberto,A. and Ciccarelli,F.D. (2008) Low duplicability and network fragility of cancer genes. *Trends Genet.*, **24**, 427–430.

34. D'Antonio,M. and Ciccarelli,F.D. (2011) Modification of gene duplicability during the evolution of protein interaction network. *PLoS Comput. Biol.*, **7**, e1002029.

35. Syed,A.S., D'Antonio,M. and Ciccarelli,F.D. (2010) Network of cancer genes: a web resource to analyze duplicability, orthology and network properties of cancer genes. *Nucleic Acids Res.*, **38**, D670–D675.

36. Anelli,V., Santoriello,C., Distel,M., Koster,R.W., Ciccarelli,F.D. and Mione,M. (2009) Global repression of cancer gene expression in a zebrafish model of melanoma is linked to epigenetic regulation. *Zebrafish*, **6**, 417–424.

37. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.

38. Santarius,T., Shipley,J., Brewer,D., Stratton,M.R. and Cooper,C.S. (2010) A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer*, **10**, 59–64.

39. Coffey,A.J., Kokocinski,F., Calafato,M.S., Scott,C.E., Palta,P., Drury,E., Joyce,C.J., Leproust,E.M., Harrow,J., Hunt,S. *et al.* (2011) The GENCODE exome: sequencing the complete human exome. *Eur. J. Hum. Genet.*, **19**, 827–831.

40. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.

41. Harrow,J., Denoeud,F., Frankish,A., Reymond,A., Chen,C.K., Chrast,J., Lagarde,J., Gilbert,J.G., Storey,R., Swarbreck,D. *et al.* (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol.*, **7(Suppl. 1)**, S41–S49.

42. Muller,J., Szklarczyk,D., Julien,P., Letunic,I., Roth,A., Kuhn,M., Powell,S., von Mering,C., Doerks,T., Jensen,L.J. *et al.* (2010) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.*, **38**, D190–D195.

43. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.

44. Stark,C., Breitkreutz,B.J., Chatr-Aryamontri,A., Boucher,L., Oughtred,R., Livstone,M.S., Nixon,J., Van Auken,K., Wang,X., Shi,X. *et al.* (2010) The BioGRID interaction database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.

45. Aranda,B., Achuthan,P., Alam-Faruque,Y., Armean,I., Bridge,A., Derow,C., Feuermann,M., Ghanbarian,A.T., Kerrien,S., Khadake,J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.

46. Ceol,A., Chatr Aryamontri,A., Licata,L., Peluso,D., Briganti,L., Perfetto,L., Castagnoli,L. and Cesareni,G. (2009) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.

47. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The database of interacting proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.

48. Inui,M., Martello,G. and Piccolo,S. (2010) MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell. Biol.*, **11**, 252–263.

49. Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.

50. Papadopoulos,G.L., Reczko,M., Simossis,V.A., Sethupathy,P. and Hatzigeorgiou,A.G. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.

51. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.

52. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

53. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.

54. Van Raamsdonk,C.D., Griewank,K.G., Crosby,M.B., Garrido,M.C., Vemula,S., Wiesner,T., Obenauf,A.C., Wackernagel,W., Green,G., Bouvier,N. *et al.* (2010) Mutations in GNA11 in uveal melanoma. *N. Engl. J. Med.*, **363**, 2191–2199.

55. Tadokoro,Y., Ema,H., Okano,M., Li,E. and Nakauchi,H. (2007) De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J. Exp. Med.*, **204**, 715–722.

56. Ley,T.J., Ding,L., Walter,M.J., McLellan,M.D., Lamprecht,T., Larson,D.E., Kandoth,C., Payton,J.E., Baty,J., Welch,J. *et al.* (2010) DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.*, **363**, 2424–2433.

57. Fabbri,M., Garzon,R., Cimmino,A., Liu,Z., Zanesi,N., Callegari,E., Liu,S., Alder,H., Costinean,S., Fernandez-Cymering,C. *et al.* (2007) MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl Acad. Sci. USA*, **104**, 15805–15810.

58. Li,B., Wang,B., Niu,L.J., Jiang,L. and Qiu,C.C. (2011) Hypermethylation of multiple tumor-related genes associated with DNMT3b up-regulation served as a biomarker for early diagnosis of esophageal squamous cell carcinoma. *Epigenetics*, **6**, 307–316.